ELSEVIER

Gene 279 (2001) 181–196

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

The human tuftelin gene: cloning and characterization ☆

Zhengkuan Mao[a], Boaz Shay[a], Maharan Hekmati[a], Eran Fermon[a], Angela Taylor[a], Leah Dafni[a], Kristiina Heikinheimo[b], Joshua Lustmann[c], Larry W. Fisher[d], Marian F. Young[d], Dan Deutsch[a,*]

[a]*Dental Research Unit, Department of Oral Biology, Institute of Dental Sciences, Hebrew University, Hadassah Faculty of Dental Medicine, Jerusalem, Israel*
[b]*Department of Oral Maxillofacial Surgery, Institute of Dentistry, University of Turku, Turku, Finland*
[c]*Department of Oral Maxillofacial Surgery, Hebrew University, Hadassah Faculty of Dental Medicine, Jerusalem, Israel*
[d]*Craniofacial and Skeletal Disease Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA*

Received 18 May 2001; received in revised form 4 September 2001; accepted 8 October 2001
Received by D. Schlessinger

Abstract

Tuftelin has been suggested to play an important role during the development and mineralization of enamel. We isolated the full-length human tuftelin cDNA using reverse transcription-polymerase chain reaction (RT-PCR) and rapid amplification of cDNA ends (5′ RACE and 3′ RACE) methods. Sequence analysis of the tuftelin cDNA revealed an open reading frame of 1170 bp encoding a 390 amino acid protein with a molecular mass of 44.3 kDa and an isoelectric point of 5.7. The human tuftelin protein shares 89 and 88% amino acid sequence identity with the bovine and mouse tuftelin, respectively. It contains a coiled-coil region, recently reported to be involved with tuftelin self-assembly and with the interaction of tuftelin with TIP39 (a novel tuftelin interacting protein). Detailed DNA analysis of the cloned genomic DNA revealed that the human tuftelin gene contains 13 exons and is larger than 26 kb. Two alternatively spliced tuftelin mRNA transcripts have now been identified in the human tooth bud, one lacking exon 2, and the other lacking exon 2 and exon 3. Primer extension analysis, corroborated by RT-PCR and DNA sequencing, revealed multiple transcription initiation sites. The cloned 1.6 kb promoter region contained several GC boxes and several transcription factor binding sites such as those for activator protein 1 and stimulatory protein 1. Our blast search of the human and mouse expressed sequence tag data bases, as well as our RT-PCR and DNA sequencing results, and a previous study using Northern blot analysis revealed that tuftelin cDNA sequences are also expressed in normal and cancerous non-mineralizing soft tissues, suggesting that tuftelin has a universal function. We have now identified and characterized different alternatively spliced mouse tuftelin mRNAs in several non-mineralizing tissues. These results provide an important baseline for future understanding of the biological role of tuftelin.

Keywords: Odontogenesis; Enamel biomineralization; Alternative mRNA splicing; Promoter; Non-mineralizing tissues

1. Introduction

Tuftelin is an acidic protein first cloned and sequenced (Deutsch et al., 1989, 1991a) from a bovine ameloblast-enriched cDNA expression library. It is found in developing and mature extracellular enamel, a unique and highly mineralized (96% mineral by tissue weight) ectodermal tissue covering vertebrate teeth. The enamel is formed within a unique extracellular organic matrix, composed mainly of proteins which are synthesized and secreted by specialized cells of the enamel organ, the ameloblasts.

Immunohistochemistry and Western blot analysis suggested that tuftelin is synthesized and secreted by the ameloblast into the enamel matrix and concentrates at the dentin–enamel junction (DEJ) region (Deutsch et al., 1991a, 1995, 1998; Zeichner-David et al., 1995, 1997; Paine et al., 2000) where enamel mineralization commences. In mature enamel, it appears as part of the tuft proteins (Deutsch et al., 1991a). There is also evidence that at very early stages of odontogenesis, the preodontoblasts that eventually form the underlying dentin also express tuftelin in a transient manner (Diekwisch et al., 1997; Zeichner-David et al., 1997).

Abbreviations: aa, amino acid(s); AP1, activator protein 1; CKII, casein kinase II; DEJ, dentin–enamel junction; E(13), embryonic day (13); ESTs, expressed sequence tags; kDa, kilodalton(s); pI, isoelectric point; PKC, protein kinase C; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-polymerase chain reaction; SP1, stimulatory protein 1

☆ GenBank™ Accession numbers: AF254260, AF254850, AF254851, AF254852, AF254854, AF254855, AF254856, AF254857, AF254858, AF254859, AF254860, AH009496.

* Corresponding author. Dental Research Unit, Department of Oral Biology, Institute of Dental Sciences, Hebrew University, Hadassah Faculty of Dental Medicine, P.O. Box 12272, Jerusalem 91120, Israel. Tel.: +972-2-675-8565; fax: +972-2-675-7307.
E-mail address: dddan@cc.huji.ac.il (D. Deutsch).

0378-1119/01/$ - see front matter
PII: S0378-1119(01)00749-1

The initial expression of tuftelin mRNA in the mouse was found at embryonic day 13 (E13), while the protein was detected by immunohistochemistry at E17 (Zeichner-David et al., 1997). The timing of tuftelin protein expression (just before enamel mineralization commences), the location where it accumulates (at the region of DEJ), and its acidic nature suggest that tuftelin might be involved in the initial stages of mineralization of enamel (perhaps by nucleation) (Deutsch et al., 1991a, 1998). However, it is also expressed at the very early stages of odontogenesis (E13), when no enamel or dentin have as yet been formed, suggesting that tuftelin might have multiple roles, being also involved in the mesenchyme/ectoderm communication (Zeichner-David et al., 1997; Deutsch et al., 1998).

Table 1
Primers used for various applications as described in the text

Name	Sequence	Location	Orientation
RT-PCR of human tuftelin			
DD101[a]	GAAGATGAACGGGACGCGTAACTGGTG	21–47[a]	S
DD107	ATGAACGGGACGCGGAACTGGTG	94–116	S
DD22	GGACATTCTCAGGCTGACTC	162–181	S
DD23	CAAGTTCATCTCCTGTCAGTTCAC	211–188	AS
DD43	CTGCCCGGGTTCTCTGTCTTGGC	1202–1180	AS
DD145	TCTCCAGGCAGCTCAGGTTTCCACC	1278–1254	AS
DD97	ACCATTCCCAAGCCCCTGGCCACTCTAAG	1403–1431	S
DD124	CATACTCTGATGCCTATCTACTTC	2852–2829	AS
3′ RACE-PCR of human tuftelin			
DD46	GAGATGCATGACCGGATGGAAC	1108–1129	S
DD55	TAGGATATCCAAGCCGCCTAGC	1206–1227	S
DD97	ACCATTCCCAAGCCCCTGGCCACTCTAAG	1403–1431	S
DD139	CCTCAGGTGATCCACCCACCTCGGCTTC	2721–2748	S
AUAP	GGCCACGCGTCGACTAGTAC	Adaptor	
5′ RACE-PCR of human tuftelin			
DD25	TACACCTTAATGATCTCCTCATG	335–313	AS
DD23	CAAGTTCATCTCCTGTCAGTTCAC	211–188	AS
DD109	CCGCCTGGTCCTCTGGGTGCACGTC	151–127	AS
AAP	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG	Adaptor	
AUAP	GGCCACGCGTCGACTAGTAC	Adaptor	
Gene walking of promoter and exon 1 of human tuftelin			
DD109	CCGCCTGGTCCTCTGGGTGCACGTC	151–127	AS
DD110	CCTCTGGGTGCACGTCCACCAGGGTA	142–117	AS
AP1	GTAATACGACTCACTATAGGGC	Adaptor	
AP2	ACTATAGGGCACGCGTGGT	Adaptor	
Gene walking of exon 1 and intron 1 of human tuftelin			
DD111	TACCCTGGTGGACGTGCACCCAGA	117–140	S
DD120	GACGTGCACCCAGAGGACCAGGCGG	127–151	S
AP1	GTAATACGACTCACTATAGGGC	Adaptor	
AP2	ACTATAGGGCACGCGTGGT	Adaptor	
Gene walking of exon 13 of human tuftelin			
DD96	GCCTGTCATCCGAGTGGTGGAAACCTGAGC	1239–1268	S
DD97	ACCATTCCCAAGCCCCTGGCCACTCTAAG	1403–1431	S
AP1	GTAATACGACTCACTATAGGGC	Adaptor	
AP2	ACTATAGGGCACGCGTGGT	Adaptor	
Primer extension of human tuftelin			
DD110	CCTCTGGGTGCACGTCCACCAGGGTA	142–117	AS
DD152	CGTGAGAAAGGATGAGCCTTGAC	(−592)–(−614)	AS
Primer for RT-PCR of mouse tuftelin			
DD112	CTCATACGACCTCCGCGTGAAGATG	1–25	S
DD107	ATGAACGGGACGCGGAACTGGTG	23–45	S
DD108	TCAGGTTTCCACCACCCGGATGAG	1195–1172	AS

[a] DD101 was designed based on the sequence of bovine tuftelin cDNA (Deutsch et al., 1991a) (position 21–47 bp). It corresponds to the human tuftelin cDNA region 90–116 bp (Fig. 2). There is one base pair difference between the bovine and corresponding human tuftelin cDNA sequence.

More recently, a genomic library clone containing the bovine tuftelin gene has been isolated and characterized (Bashir et al., 1997). The gene directs the synthesis of at least four tuftelin mRNA transcripts through alternative RNA splicing. The genomic DNA sequence differed from the original reported cDNA sequence (Deutsch et al., 1991a), resulting in a change of the reading frame at the carboxy-terminus. This suggested that there might be two isoforms, A and B (the latter being the most common form found in bovine (Bashir et al., 1997), mouse, porcine (MacDougall et al., 1998) and now human (Deutsch et al., 1998) (also see Section 3)). Using an isolated and partially characterized human tuftelin clone, screened from a human genomic library, we localized the human tuftelin gene to chromosome 1q 21-31 (Deutsch et al., 1994) (see also the National Center for Biotechnolgy Information (NCBI) under Map Viewer, http://www.ncbi.nlm.nih.gov/). Northern blot analysis revealed that tuftelin is also expressed in kidney, liver, lung and testis (MacDougall et al., 1998). This important finding opened new horizons into the possible role of tuftelin in these different tissues.

The present study describes the isolation and characterization of the complete human tuftelin gene. It includes information on exon–intron organization and structure, the transcription initiation sites, the promoter structure and newly discovered alternatively spliced tuftelin mRNA transcripts, none of which have been described in the past. The new data also help to assign the start codon to exon 1, and not to exon 3 as had been assigned previously in bovine (Bashir et al., 1997). The data describe new information on the structural motifs in the human-derived tuftelin protein and its relation to tuftelin from other species.

These new data provide a baseline for better understanding of the structure and function of tuftelin in the development and mineralization of enamel, in normal and abnormal odontogenesis, and in normal and cancerous non-mineralizing tissues.

2. Materials and methods

2.1. RT-PCR of human tooth bud RNA

Tuftelin gene-specific oligonucleotide primers were synthesized in the interdepartmental equipment unit in the Hebrew University School of Medicine or by Eisenberg Bros. Ltd., Israel. The sequences of the specific human tuftelin oligonucleotide primers used are shown in Table 1. Human tooth bud total RNA (0.29 μg) (for extraction method, see Catalano-Sherman et al., 1993) was reverse-transcribed using a random primer and SuperScript II reverse transcriptase (Life Technologies, Inc.) according to instructions provided by the manufacturer in a total reaction volume of 20 μl. Single-stranded cDNA (1 μl) was amplified with tuftelin gene-specific primers (DD101/DD23, DD22/DD43, DD97/DD124, DD107/DD145) in a total reaction volume of 50 μl containing NH_4 buffer, 1.5 mM $MgCl_2$, 200 μM of dNTP, 0.2 μM of each primer and 2.5 units of BIOTAQ DNA polymerase (BIOLINE). The PCR reaction was as follows: 3 min at 94°C, followed by 35 cycles at 94°C for 30 s, 55°C for 45 s and 72°C for 2 min, followed by a final extension at 72°C for 10 min.

2.2. RT-PCR of tuftelin RNA isolated from mouse liver, kidney and lung

Total RNA of mouse liver, kidney and lung was extracted and reverse-transcribed as described above. Tuftelin cDNA was PCR-amplified using DD112 and DD108 as follows: 94°C for 2 min, followed by 35 cycles at 94°C for 30 s, 55°C for 45 s and 72°C for 2 min, followed by a final extension at 72°C for 10 min. Because the first PCR produced weak bands, a second PCR was carried out using nested primers, DD107 and DD108, under the same conditions as above.

2.3. 3′ rapid amplification of human tuftelin cDNA ends

The 3′ end of the human tuftelin transcript was cloned using the 3′ RACE method (Life Technologies, Inc.) (Frohman et al., 1988). Human tooth bud total RNA (0.29 μg) was reverse-transcribed using an oligo-dT adapter primer and SuperScript II reverse transcriptase (Life Technologies, Inc.). A total of 1 μl of the reaction product was amplified by PCR using a human tuftelin gene-specific primer DD46 and an adapter-specific primer AUAP. Nested PCR was performed with 1 μl of the reaction mixture obtained from the previous round of PCR as a template using the nested primer DD55 and an adaptor primer AUAP. Another 3′ RACE experiment was performed using DD97 and the nested primer DD139.

2.4. 5′ rapid amplification of the human tuftelin cDNA ends

The 5′ end of the human tuftelin transcript was cloned using the 5′ RACE method (Life Technologies, Inc.) (Frohman et al., 1988). Briefly, single-stranded cDNA was reverse-transcribed as before using the antisense tuftelin-specific primer DD25. After removal of the RNA template, using RNase H, an anchor priming site at the 3′ end of the single-stranded cDNA was created using dCTP and terminal deoxynucleotidyl transferase, which added a homopolymeric tail. Tailed cDNA was amplified by PCR using a tuftelin-specific nested primer DD23 and the adapter primer AAP. Nested PCR was performed using DD109 and the adapter primer AUAP.

2.5. Cloning and sequencing of the human tuftelin cDNA

The PCR products of the 5′ RACE, 3′ RACE and RT-PCR were cloned into the pCR 2.1 vector (TA Cloning Kit, Invitrogen). The ligation products were transformed into One Shot competent cells (Invitrogen). Recombinant plasmids were digested by *Eco*RI restriction enzyme and the positive clones determined by Southern blot hybridization.

The plasmids of the positive clones were purified by using QIAGEN Plasmid Mini or Midi Kits. DNA was sequenced by means of an automated sequencer (ABI prizm 377; Perkin-Elmer, USA; DNA Sequencing Center, Silberman Institute of Life Sciences, Hebrew University) with universal primers or a series of specific oligonucleotides. The clones were sequenced on both strands several times.

2.6. Cloning and sequencing the human tuftelin genomic DNA

A clone containing almost the entire human tuftelin gene has been isolated from a lambda fix (Stratagene) human genomic library, using ^{32}P-labeled bovine tuftelin cDNA as a probe (Deutsch et al., 1994). The insert size of this clone is greater than 20 kb. The clone insert DNA was digested using several restriction enzymes and the different DNA fragments were subcloned into pBluescript II KS (Stratagene) plasmid and characterized by Southern blot analysis according to standard procedures (Sambrook et al., 1989). Using gene-specific sequencing primers, the exons and exon–intron junctions were sequenced. DNA sequence analysis showed that this genomic clone contained human tuftelin gene exons 2–12, but did not contain exon 1 and exon 13.

2.7. 5′ and 3′ genomic DNA walking

The 5′ and 3′ flanking regions of the human tuftelin gene, which were not contained within the human tuftelin genomic clone, were cloned by the PCR-based gene-walking method (Human Genome Walker Kit, Clontech). Amplification of the unknown 5′ and 3′ regions of the human tuftelin gene was performed by using the five different human genome walker genomic DNA libraries with two rounds of PCR using the adapter primers AP1, AP2 and the two gene-specific primers (listed in the Table 1). Exon 1 and its 5′ flanking region were amplified with a primer DD109 and the nested primer DD110. The DNA region of exon 1 and intron 1 was obtained by using the primer DD111 and the nested DD120. Exon 13 and its 3′ flanking region were cloned with DD96 and DD97. Primary amplification of gene walking was performed using the adapter primer AP1 and tuftelin gene-specific primers. The PCR reaction was as follows: seven cycles at 94°C for 25 s and 72°C for 4 min, followed by 35 cycles at 94°C for 25 s and 67°C for 4 min, with a final extension at 67°C for 4 min. A total of 1 μl of primary PCR product was used as the template for the secondary PCR using the nested, adapter primer AP2 and a human tuftelin-specific nested primer. The same protocol was used as described above, with the exception of the first and second steps, which consisted of five and 22 cycles, respectively. The PCR products were subcloned into the pCR 2.1 vector (Invitrogen) and sequenced using gene-specific oligonucleotide primers.

2.8. Identification of exons, exon–intron junctions and determination of the intron sizes in the human tuftelin gene

Exon–intron junctions were established by comparison of the genomic DNA and the cDNA sequences. Intron sequences were amplified using human genomic DNA (Promega) and tuftelin flanking exon-specific primers. Accu Taq LA DNA Polymerase Mix (Sigma) was used to amplify genomic DNA fragments of more than 2 kb. The sizes of the introns were determined by measuring the sizes of PCR products using agarose gels or by DNA sequencing.

2.9. Determination of human tuftelin transcription initiation sites using primer extension analysis

In order to establish the transcription initiation site, primer extension analyses were performed using two antisense oligonucleotide primers: (a) DD110 made to a region near to the ATG start codon and located downstream 21–46 bp from the ATG start codon, i.e. 117–142 bp downstream from +1; and (b) DD152 made to a region towards the 5′ end of the human lung tuftelin cDNA obtained by RT-PCR and located 707–685 bp upstream to the ATG start codon, i.e. −614 to −592 from +1. Both these primers were 5′ end-labeled with [γ-^{32}P]ATP (Amersham) using T4 polynucleotide kinase. Each labeled primer and 5 μg of human lung total RNA (Stratagene) were annealed at 70°C for 10 min and then slowly cooled to room temperature. The annealed primer was extended at 42°C for 1 h using SuperScript II reverse transcriptase (Life Technologies, Inc.) and the reaction products were electrophoresed on an 8% polyacrylamide gel. The DNA sequencing marker (ACGT), produced by using the *fmol*® DNA cycle sequencing system (Promega), represents the sequence of the 5′ region of the human tuftelin gene with the same primer as that used for primer extension. To corroborate the results of primer extension analysis, RT-PCR and DNA sequences were carried out.

2.10. Nucleotide and amino acid sequence analysis

Nucleotide analysis was carried out using the GCG program. The molecular weight (MW), the theoretical isoelectric point (pI), and amino acid composition were analyzed and calculated by using the ProtParam Program (http://www.expasy.ch/tools/protparam.html). Multiple sequence alignment of tuftelin protein was analyzed by the CLUSTAL W Program (http://www.clustalw.genome.ad.jp/). Potential post-translational modifications such as N-glycosylation sites, protein kinase C and casein kinase II phosphorylation sites, amidation and N-myristoylation sites were identified using the PROSCAN program (http://pbil.ibcp.fr/) and O-glycosylation sites using the NetOGlyc 2.0 program (http://www.cbs.dtu.dk/services/NetOGlyc/). Blast searches of the human expressed sequence tagged (EST) database and mouse EST database were carried

Fig. 1. Strategy for the cloning of the human tuftelin cDNA. The previously described exons (Deutsch et al., 1994) are shown in hatched boxes and unknown exons are shown in white boxes. The fragments with the arrows were obtained with RACE-PCR. The fragments with black dots were obtained using RT-PCR (see Section 3.1).

out in the NCBI (http://www.ncbi.nlm.nih.gov/blast/). The promoter prediction software (http://www.fruitfly.org/seq_tools/promoter.hfml) was used to predict the transcription initiation site. Putative transcription factor binding sites were searched by MatInspector 2.2 software (http://genomatix.gsf.de/cgi-bin/matinspector/matinspector.pl).

3. Results

3.1. Isolation and characterization of the human tuftelin cDNA

In a previous study, we reported the screening of a lambda phage human genomic library and isolation of a genomic clone containing the human tuftelin gene (Deutsch et al., 1994). The cloned genomic DNA was initially sequenced using subcloned fragments in the KS plasmid and double-stranded sequencing of the plasmid with T3, T7 and some synthesized bovine tuftelin oligonucleotide primers. Five exons of the human tuftelin gene (exons 2, 3, 5, 7 and 12) were originally identified and sequenced (Deutsch et al., 1994). These sequences, and a partial published segment of exon 13 (the first 64 bp) (Bashir et al., 1998), were used to design the primers for 5′ RACE, 3′ RACE and RT-PCR of the human tuftelin cDNA. Our strategy for cloning the full-length human tuftelin cDNA is summarized in Fig. 1. The composite nucleotide sequence of the human tuftelin cDNA generated by RT-PCR, 5′ RACE and 3′ RACE is shown in Fig. 2. The full-length human tuftelin cDNA is at least 3139 nucleotides in length

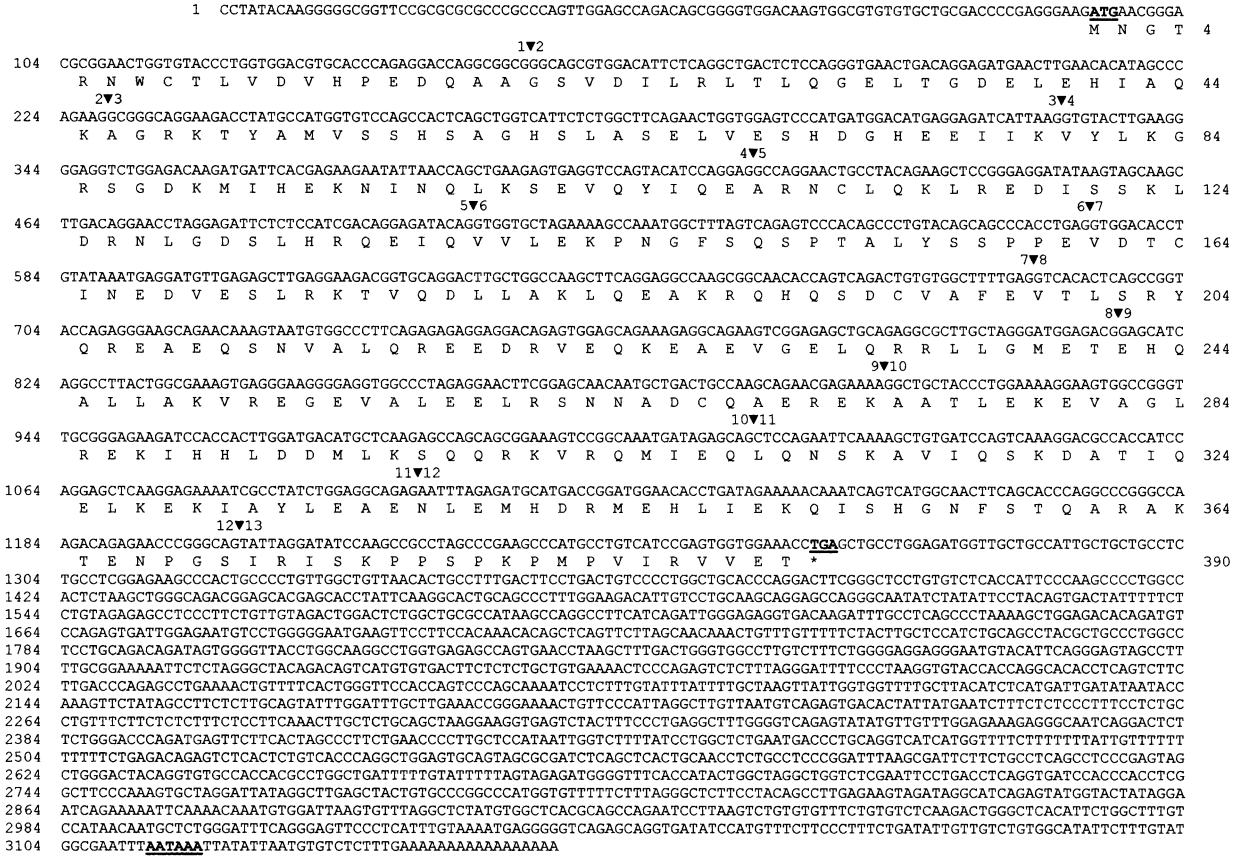

Fig. 2. Nucleotide sequence of the human tuftelin cDNA and deduced amino acid sequence. The number of the first nucleotide and the final amino acid in each row are indicated on the left and right side, respectively. Division of the sequence into exons is indicated (▼). The start codon (ATG), the stop codon (TGA) and a polyadenylation consensus sequence (AATAAA) are underlined (see Section 3.1).

(referring to the transcription initiation site (+1) (see Fig. 7)), exclusive of the poly(A)$^+$ tail. Two in-frame ATG codons were located at nucleotides 94–96 and 250–252 bp, respectively. The nucleotides between these two ATG codons are highly conserved in the human, bovine and mouse tuftelin. The first ATG (94–96 bp) was assigned as the translation start codon (see Section 4), which was followed by an open reading frame of 1170 bp encoding a 390 amino acid protein with a molecular mass of 44.3 kDa and a pI of 5.7. The human tuftelin translation stop codon TGA is at 1264–1266 bp, after which a long untranslated region of 1873 nucleotides begins. There is only one polyadenylation signal, AATAAA, which is located at 3113–3118 bp, 21 nucleotides upstream of the poly(A)$^+$ tail.

3.2. Determination of human tuftelin transcription initiation sites

The 5′ RACE results identified a human tooth bud tuftelin cDNA transcript that extended to 62 bp upstream to the ATG start codon (see Fig. 7). Promoter prediction software predicted, with a perfect score, a transcription initiation site at the exact location identified by 5′ RACE. A TATA box and a reverse CAAT box were identified 29 and 123 bp upstream to the 5′ RACE predicted initiation site, respectively.

RT-PCR, using human lung total RNA and primers upstream to the ATG start codon (1700–1672, 1423–1401, and 1013–990 bp, respectively), together with a common downstream reverse primer corresponding to a region in exon 2 (92–115 bp downstream to ATG, corresponding to tuftelin cDNA sequence 118–211 (see Fig. 2)), did not produce DNA bands. PCR products with the expected sizes were, however, obtained using upstream primers located at 707–682, 363–338, 167–143, 105–83, and 61–41 bp upstream to the ATG start codon, and the common reverse primer (cDNA region 188–211 bp) (data not shown). Sequencing of the largest RT-PCR product (i.e. that generated using an upstream primer located at 707–682 bp from the ATG start codon and the common downstream reverse primer) revealed the exact sequence of tuftelin corresponding to this region. Since the reverse downstream primer corresponds to exon 2, this excluded the possibility that such PCR amplification was of genomic DNA (5′ to exon 2 there is >4.4 kb intron in genomic DNA). These results suggested that between 1013 and 707 bp, upstream to the start codon ATG, lies the furthest upstream initiation transcription site.

Primer extension analyses, using human lung total RNA and one of the two reverse primers, either DD110 or DD152, revealed multiple initiation sites at 93 and 110 bp and at 758, 779, and 796 bp upstream to the start codon ATG (Fig. 3). The primer extension results are consistent with the RT-PCR results and suggest the existence of multi-transcription initiation sites. Further detailed studies will be carried out in future to characterize the untranslated human tuftelin cDNA region and possible additional transcription initiation sites.

There is still no definite evidence, at least from the primer extension results, to support a transcription initiation site at 62 bp upstream to ATG as suggested by the 5′ RACE results.

We have designated the most proximally located transcription initiation site (C), 93 bp upstream to the translation initiation site ATG, as +1.

3.3. The structure of the human tuftelin gene

The structure of the human tuftelin gene was initially studied by isolating a clone from the lambda fix human genomic library followed by DNA sequencing (Deutsch et al., 1994). As mentioned previously, this clone was subsequently proved to contain exons 2–12 but lacks exons 1 and 13. To establish the sequences of the genomic regions corresponding to exons 1 and 13, and the corresponding flanking intron 1 and 12 regions, the PCR-based gene-walking method was used. The corresponding exons and exon–

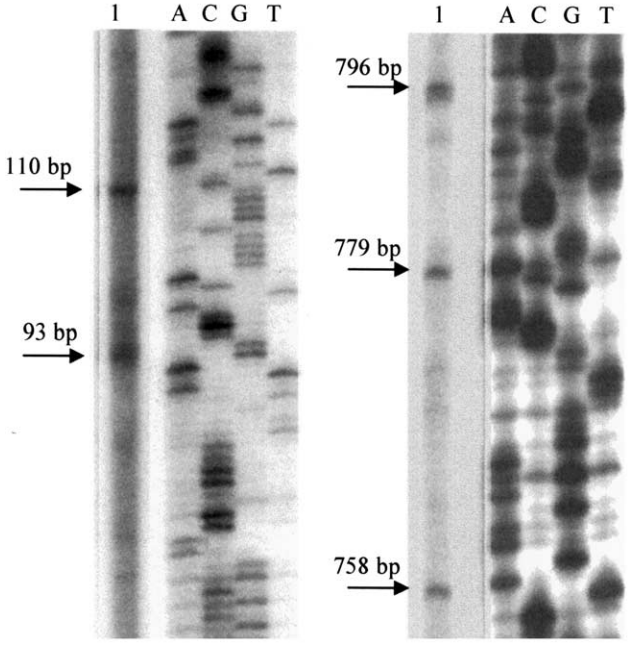

Fig. 3. Determination of transcription initiation sites by primer extension. Total RNA from human lung was used for mapping the transcription initiation sites with two oligonucleotides: (a) DD110 made to the region near to the ATG start codon and located downstream 21–46 bp from the ATG start codon, i.e. 117–142 bp downstream from +1 (left side); and (b) DD152 made to a region towards the 5′ end of the human lung tuftelin cDNA obtained by RT-PCR and located at 707–685 bp upstream to the ATG start codon, i.e. −614 to −592 from +1 (right side). Each ^{32}P-end-labeled oligonucleotide was hybridized with 5 μg of the total RNA and reverse-transcribed using SuperScript II RT (Life Technologies, Inc.). The primer extension products were electrophoresed on an 8% sequencing gel along with sequencing reaction products used as size markers. Extension products corresponding to the transcription initiation sites at 93 and 110 bp and at 758, 779, and 796 bp upstream to the translation start codon (ATG) are indicated. The DNA sequencing marker (ACGT) represents the sequence of the 5′ region of the human tuftelin gene determined with the same primer as that used for primer extension (see Section 3.2). Notice that the deduced hybridization sequence is complementary to the human tuftelin cDNA sequence.

Table 2
Exon–intron organization of the human tuftelin gene[a]

Exon	Exon size (bp)	Sequences at exon–intron junctions				Amino acid interrupted
		5′-splice donor	Intron (kb)[b]	(kb)[c]	3′-splice acceptor	
1	153	GCGGCG**gt**aaga	> 3.0	≥ 4.4	tatt**ag**GGCAGC	Ala-/-Gly
2	75	CAGAAG**gt**actg	0.419	0.419	ggcc**ag**GCGGGC	Lys-/-Ala
3	102	ATTAAG**gt**aagc	1.3	1.217	tttc**ag**GTGTAC	Lys-/-Val
4	87	CAGGAG**gt**gggc	0.534	0.534	tccc**ag**GCCAGG	Glu-/-Ala
5	90	ATACAG**gt**aata	1.6	1.548	gaac**ag**GTGGTG	Gln-/-Val
6	66	CCTGAG**gt**aggt	3.5	3.428	ctgt**ag**GTGGAC	Glu-/-Val
7	114	TTTGAG**gt**gaga	4.5	4.499	tttt**ag**GTCACA	Glu-/-Val
8	129	GAGACG**gt**aacc	0.512	0.512	accc**ag**GAGCAT	Thr-/-Glu
9	95	AGAAAA**gt**aagg	3.7	3.715	cccc**ag**GGCTGC	Glu-Ly/s-Ala
10	106	GAGCAG**gt**aagt	0.822	0.822	ttgc**ag**CTCCAG	Gln-/-Leu
11	84	GCAGAG**gt**gtgt	1.1	≥ 1.133	ttac**ag**AATTTA	Glu-/-Asn
12	101	GGGCAG**gt**gagt	0.601	0.602	tttc**ag**TATTAG	Gly-Se/r-Ile
13	1941					

[a] The sizes of exons and introns are indicated. Exon sequences are in uppercase. Intron sequences are in lowercase. The size of exon 1 includes a 93 bp untranslated region and the size of exon 13 includes a 1877 bp untranslated region (excluding TGA).

[b] Intron sequences were amplified using human genomic DNA and tuftelin flanking exon-specific primers. The sizes of introns were determined by the sizes of PCR products on agarose gel (3, 5, 6, 7, 9, 11) or by DNA sequencing (intron 2, 4, 8, 10, 12). The exact size of intron 1 has not been determined due to failure of amplifying the entire region of intron 1.

[c] Intron sizes were calculated based on the sequence of the Human Genome Project data (GenBank Accession number: AL365436). The exact sizes of intron 1 and intron 11 are not known because there are gaps between the fragments containing exon 1 and exon 2 and between fragments containing exon 11 and exon 12.

intron boundary regions in the DNA of the human tuftelin genomic clone (including the DNA regions corresponding to exons 1 and 13) and the human tuftelin cDNA sequence were subsequently compared (Table 2). The results showed that the human tuftelin gene spans >26 kb of the genomic DNA and contains 13 exons (Fig. 4). The sizes of the exons varied from 66 bp (exon 6) to 1941 bp (exon 13). Exon 1 contains a 93 bp untranslated region and the start codon (ATG), and exon 13 contains the stop codon (TGA) and a 1877 bp untranslated region at the 3′ end (the human tuftelin cDNA sequence lacked TTTT (positioned between 2497 and 2498) at the untranslated 3′ region as compared with the corresponding genomic sequence). Note that exons 9, 10, 12 and 13 either start (exon 10, exon 13) or end (exon 9, exon 12) in the middle of a codon. The sequences of the exon–intron junctions of the human tuftelin gene conform with the consensus sequence ag-exon-gt. The sizes of the introns of the human tuftelin gene were obtained by DNA sequencing or by long range PCR. The smallest intron (intron 2) was found to be 0.4 kb in size, and the longest introns were intron 7 (4.5 kb) and intron 1. The precise length of intron 1 is not clear. According to our partial PCR sequencing results it is >3 kb and from the human genomic project it is ≥4.4 kb.

3.4. Identification of alternatively spliced transcripts of the human tooth bud tuftelin cDNA

To determine whether alternative splicing of human tuftelin mRNA occurs, we used RT-PCR to amplify human tooth

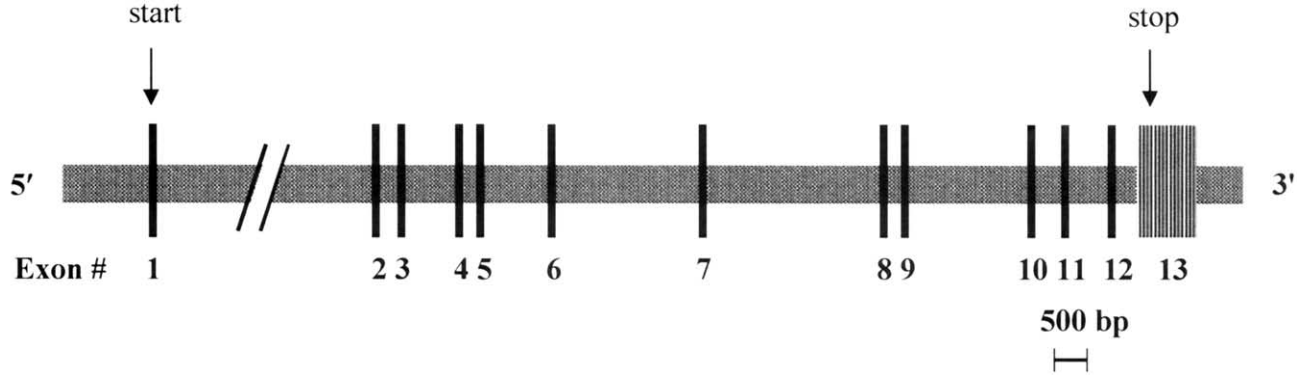

Fig. 4. The structure of the human tuftelin gene. Exons are represented by vertical bars and introns by a horizontal line. The exact distance between exon 1 and exon 2 has not been determined due to failure to isolate or amplify the entire region of intron 1. The positions of the translation start codon (ATG) in exon 1 and of the termination codon (TGA) in exon 13 are indicated (see Section 3.3).

bud tuftelin RNA using N-terminal coding oligonucleotide primer DD107, and the C-terminal coding oligonucleotide primer DD145. RT-PCR products were cloned into the pCR 2.1 vector and sequenced. The sequence analysis revealed the existence of at least two alternatively spliced mRNA transcripts, one which lacked exon 2 and the other which lacked both exon 2 and exon 3 (Fig. 5A). The alternatively spliced human tooth bud mRNA transcript, which lacks exon 2, codes for a predicted 365 amino acid protein, with a MW of 41,546 Da and a pI of 5.9 (Fig. 5B). The alternatively spliced human tooth bud mRNA transcript, which lacks exons 2 and 3, codes for a predicted protein of 331 amino acids with a MW of 37,929 Da and a pI of 5.8 (Fig. 5B).

3.5. Identification of alternatively spliced transcripts of the mouse kidney, liver and lung tuftelin cDNA

Using RT-PCR of the mouse kidney, liver and lung tuftelin RNA, followed by cDNA sequencing, a number of different alternatively spliced tuftelin mRNA transcripts were identified. A tuftelin mRNA transcript coding for all 13 exons was identified in mouse lung and liver. A spliced tuftelin mRNA transcript lacking exon 2 was identified in mouse kidney. In addition, a spliced tuftelin mRNA transcript, which lacks exon 2, the 3′ region of exon 4, exons 5–11 and the 5′ region of exon 12, was also identified in mouse liver. The mouse mRNA transcript, which contains all 13 exons (see liver and lung, Fig. 6A), codes for a predicted protein of 390 amino acids with a MW of 44.6 kDa and an acidic pI of 5.8. The mouse alternatively spliced mRNA transcript, which lacks exon 2, codes for a predicted protein of 365 amino acids with a MW of 41,828 Da and an acidic pI of 6.0 (Fig. 6B). The mouse alternatively spliced mRNA transcript, which lacks exon 2, part of exon 4 (bases 308–346), exons 5–11 and part of exon 12 (bases 1031–1090), codes for a predicted protein of 104 amino acids with a MW of 11,468 Da and a basic pI of 9.2 (Fig. 6B). These alternatively spliced mRNA-derived proteins have different MWs, amino acid compositions and pI characteristics.

3.6. DNA sequence analysis of the putative human tuftelin promoter region

To begin elucidating the molecular mechanisms controlling tuftelin expression, the structure of the putative promoter region of the tuftelin gene was established. A DNA fragment containing a 1.6 kb region upstream from the transcription initiation site of the tuftelin gene was identified using PCR-based gene walking and subsequently sequenced. DNA sequence analysis revealed that the tuftelin promoter contains numerous putative binding sites for known transcription factors and several GC boxes, which are important for transcription of the TATA-less promoters. It also contains several transcription factor binding sites

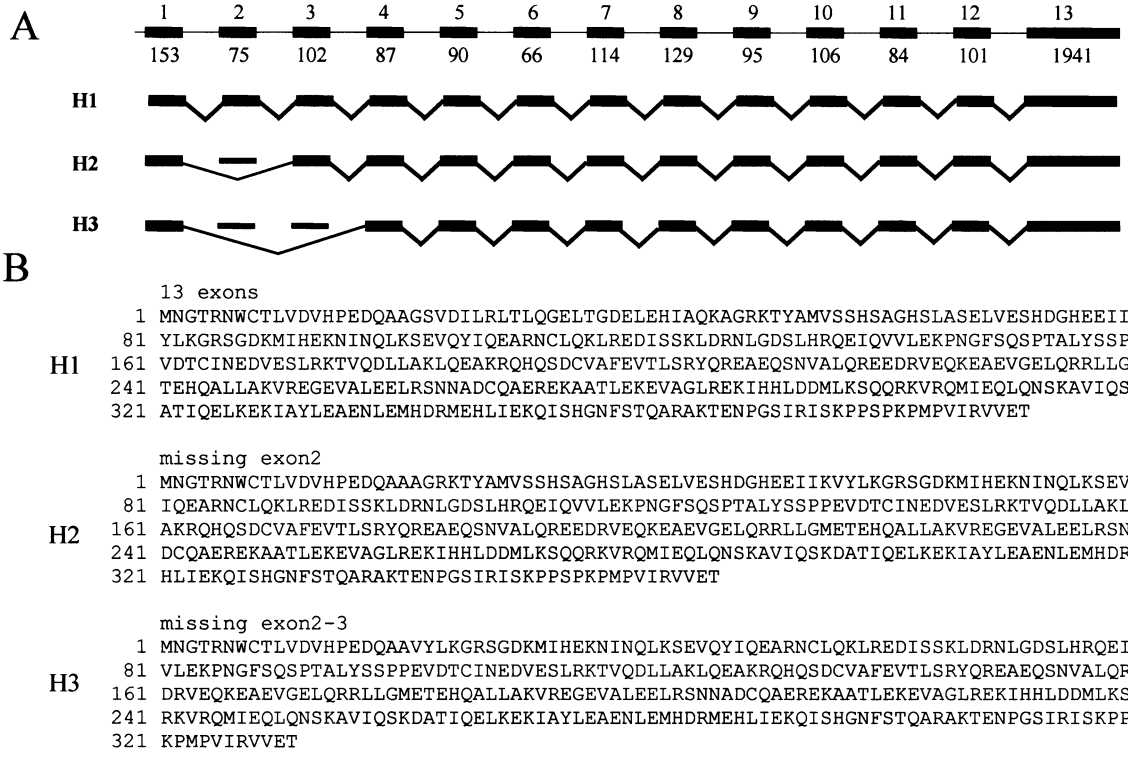

Fig. 5. (A) Alternative splicing characterizing the human tuftelin gene. (B) The corresponding derived protein sequences. The intron–exon structure of the human tuftelin gene is presented at the top of the splicing diagram; the thin lines represent introns and the thick bars exons. The exons are numbered (1–13) above the bars and the number of bases of each exon is written below. A separate non-connecting thin line represents exons removed by splicing (see Section 3.4).

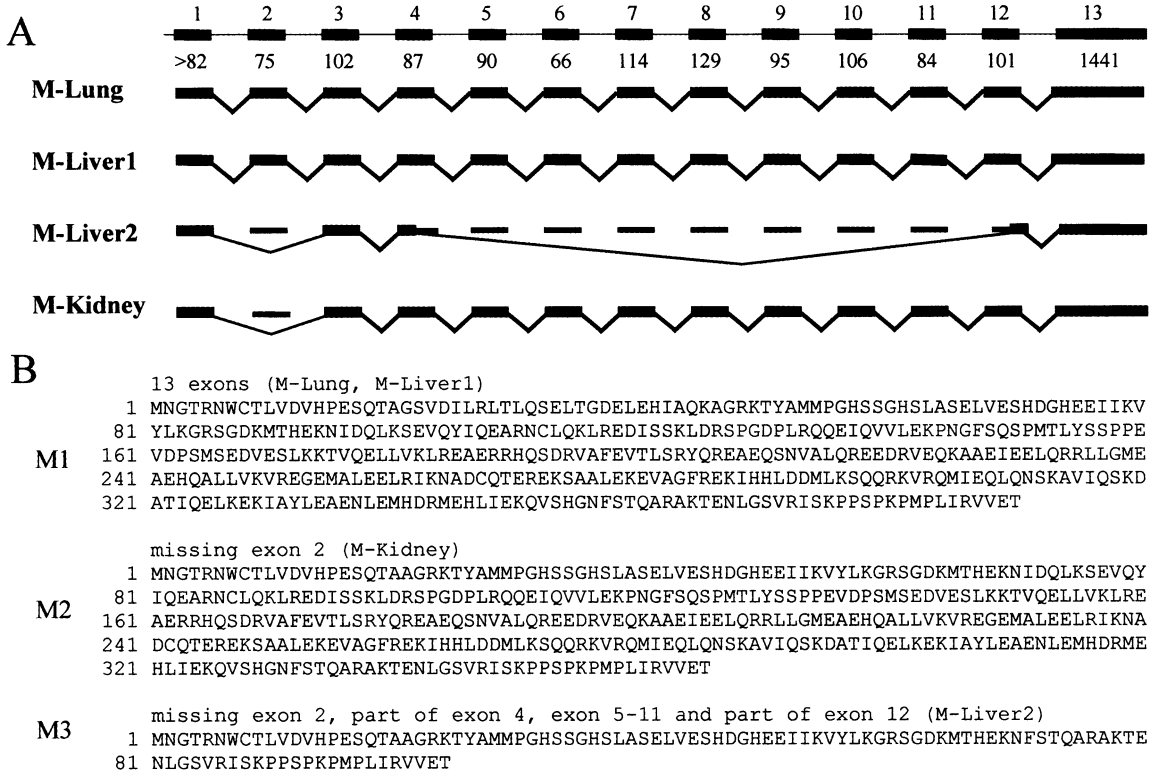

Fig. 6. (A) Characterizing alternative splicing in the putative mouse tuftelin gene. (B) The corresponding derived protein sequences. The putative structure of the mouse tuftelin gene is derived from comparing the sequence of the mouse tuftelin cDNA (GenBank Accession number: AF047704) and the human and bovine tuftelin gene structures. The exact size of exon 1 is not clear. No information is available on the transcription initiation site of the mouse tuftelin gene. The intron–exon structure of the mouse tuftelin gene is presented above the splicing diagram; the thin lines represent introns and the thick bars exons. The exons are numbered (1–13) above the bars and the number of bases of each exon is written below. A separate non-connecting thin line represents exons removed by splicing (see Section 3.5).

such as those for AP1 and SP1, which are well-known to be important for controlling gene transcription (Fig. 7).

3.7. Amino acid sequence analysis

The cDNA-derived human tuftelin protein is composed of 390 amino acids, as are the bovine and mouse tuftelins. It has a MW of 44,264 Da and a pI of 5.7. The corresponding MWs and pIs of bovine and mouse tuftelins are 44.3 kDa and 5.7, and 44.6 kDa and 5.8, respectively. The three proteins have similar amino acid compositions (Table 3). A comparison between the amino acid sequence of human tuftelin with those of bovine tuftelin and mouse tuftelin shows a high identity of 89 and 88%, respectively (Fig. 8). These results show that tuftelin is highly conserved within vertebrate species. Computer analysis of predicted post-translation modification revealed that the human tuftelin potentially possesses two N-glycosylation sites, five O-glycosylation sites, ten phosphorylation sites, one amidation site and one N-myristoylation site (Table 4). These modifications and their specific locations are conserved almost entirely in human, bovine and mouse tuftelin. Of the five cysteines present in the human tuftelin, three cysteines (residues 8, 112, 267) are also conserved in the bovine and mouse tuftelin proteins (Fig. 8), suggesting that at least one potential disulfide bridge may be part of the tuftelin structure. Analysis with a coiled-coil prediction program, Paircoil, revealed that large segments (196–252, 253–352 residues) of the tuftelin protein could adapt a coiled-coil structure with a probability of 1.000. It is interesting that an EF-hand calcium binding domain consensus sequence was identified in the predicted human, bovine and mouse tuftelin proteins (with 91, 94 and 93% similarity, respectively). It is located towards the N-terminal region of the tuftelin protein (residues 125–137).

4. Discussion

The present study is the first to describe the structure of the human tuftelin gene. It includes the structure of the putative promoter region, the complete cDNA structure, the exon–intron organization and structure, the different alternatively spliced mRNA transcripts, the primary structure of the deduced human tuftelin protein, the deduced protein characteristics, and its possible functions. It also describes numerous alternatively spliced mRNA mouse tuftelin transcripts identified in non-mineralizing tissues of kidney, liver and lung.

The human tuftelin cDNA sequence contains two poten-

tial in-frame start codons (ATG at positions 94–96 and 250–252, respectively). The second start codon (position 250–252) is surrounded by a homologous Kozak sequence (GCCATGG) (the consensus sequence for a eukaryotic initiation site). Originally, when the cDNA of tooth bud bovine tuftelin was first deciphered, and no other cDNA or genomic data were available, this second ATG codon was assigned as the start codon, recognizing, however, that other potential 5′ start codons were possible (Deutsch et al., 1989). When the bovine tuftelin gene was deciphered, this second potential start codon present in exon 3 (and not the first start codon present in exon 1) was assigned as the start codon (Bashir et al., 1997).

More recently, when the structure of the mouse tuftelin cDNA was established, it was argued (MacDougall et al., 1998) that because the most 5′ ATG is conserved between mouse and bovine, has an adenosine in the −3 position, is encoded by exon 1 (deduced from the bovine gene structure), and because there is a very high conservation of the derived amino acid sequences, starting at the first ATG, the start codon should be reassigned to the first ATG. In the present study involving the human tuftelin gene, the first ATG (exon 1, positions 94–96) was indeed assigned as the start codon. In addition to the high conservation of the amino acid sequence translated from the first ATG (exon 1) to the second ATG (exon 3) between species (the identity between the human tuftelin amino acid sequence corresponding to this latter region, and the corresponding bovine and mouse tuftelin regions is 90 and 94%, respectively), we have identified an alternatively spliced human tuftelin cDNA transcript that does not contain exon 2 and exon 3. Since exon 3 contains the second potential ATG start codon, and normally alternative splicing produces mRNA transcripts that are translated, we concluded that the start codon is indeed the ATG in exon 1. This does not exclude the possibility, however, that tuftelin might have multiple start codons and that there may be cases where the second codon ATG located at exon 3 acts as a start codon, both in tooth bud tuftelin and in tuftelin of other tissues.

In the present study, in addition (as mentioned above) to the full-length human tuftelin tooth bud mRNA transcript coding for all 13 exons, an alternative spliced tuftelin mRNA transcript missing exon 2 was found. We have also found for the first time an alternatively spliced tooth bud tuftelin mRNA transcript which lacks exons 2 and 3. All these alternative RNA splicings are of the O-type occurring between codons. Deletion of exon 2 or exons 2 and 3 would not change the tuftelin cDNA open reading frame. Alternative mRNA splicing of tuftelin has been reported in bovine tooth bud and in mouse kidney. Using Northern blot analysis, Deutsch et al. (1991a) detected the presence of two bovine tuftelin mRNA species (2.7 and 1.0 kb). Later, with the cloning of the bovine tuftelin gene (Bashir et al., 1997), four tuftelin mRNA transcripts were identified; one cDNA clone contained all 13 exons and three clones lacked DNA fragments corresponding either to exon 2, exon 5 or exon 6. MacDougall et al. (1998) reported the presence of a tuftelin mRNA transcript lacking exon 2 in mouse kidney.

The human tuftelin cDNA structure and the derived protein are of the B form identified in bovine (Bashir et al., 1997) and in mouse (MacDougall et al., 1998) (for review, see Deutsch et al., 1998).

The human tuftelin gene is a single copy gene and spans over 26 kb. It is composed of 13 exons and 12 introns. The

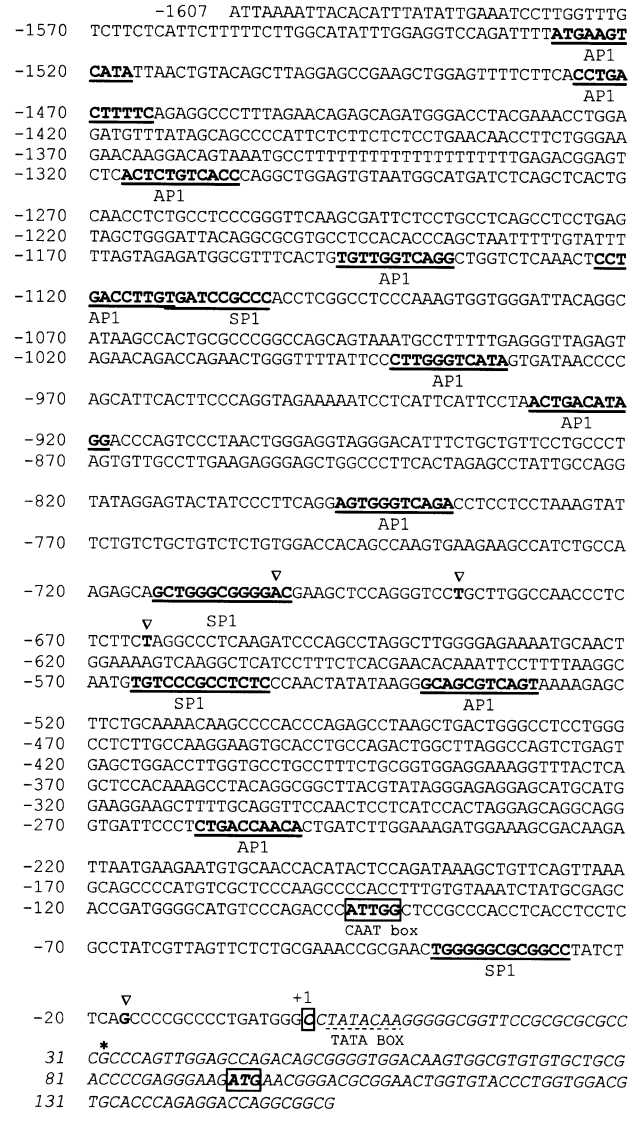

Fig. 7. Nucleotide sequence of the 5′ flanking region of the human tuftelin gene. Multiple transcription initiation sites were identified in the human lung tuftelin and their positions are marked by (▽). The most proximal transcription initiation site located 93 bp upstream to the ATG start codon is designated by (+1). The transcription initiation site identified by 5′ RACE in human tooth bud tuftelin and located 62 bp upstream to the ATG start codon is marked by (*) and the related TATA box and reversed CAAT box preceding it are underlined and boxed, respectively. Transcription factors AP1 and SP1 binding sites, which are well-known to be important for controlling gene transcription, are marked in bold and are underlined. The first exon is in italics. The translation start codon (ATG) is indicated in bold and is boxed (see Section 3.6).

Table 3
Comparison of amino acid compositions, MWs and pIs between human, bovine and mouse tuftelin

Amino acid[a]	Human tuftelin		Bovine tuftelin		Mouse tuftelin	
	Number	R/1000	Number	R/1000	Number	R/1000
Alanine	29	74	24	62	25	64
Cysteine	5	13	4	10	3	8
Aspartic	19	49	19	49	18	46
Glutamic	48	123	47	121	50	128
Phenylalanine	3	8	2	5	4	10
Glycine	17	44	17	44	16	41
Histidine	15	38	13	33	14	36
Isoleucine	20	51	20	51	18	46
Lysine	28	72	31	79	29	74
Leucine	39	100	38	97	39	100
Methionine	9	23	9	23	13	33
Asparagine	15	38	15	38	11	28
Proline	11	28	10	26	14	36
Glutamine	30	77	35	90	29	74
Arginine	26	67	23	59	28	72
Serine	29	74	31	79	32	82
Threonine	15	38	19	49	15	38
Valine	25	64	25	64	25	64
Tryptophan	1	3	1	3	1	3
Tyrosine	6	15	7	18	6	15
Asx (Asp + Asn)	(34)	(87)	(34)	(87)	(29)	(74)
Glx (Glu + Gln)	(78)	(200)	(82)	(211)	(79)	(202)
Total residues	390	1000	390	1000	390	1000
MW (kDa)	44.3		44.3		44.6	
pI	5.7		5.7		5.8	

[a] The listed values of human tuftelin, bovine tuftelin (Bashir et al., 1997) (GenBank Accession number: AAC84147) and mouse tuftelin (MacDougall et al., 1998) (GenBank Accession number: AAC04577) were obtained by the ProtParam tool (http://www.expasy.ch/tools/protparam.html).

protein coding sequence of the human tuftelin is distributed in all 13 exons. A comparison of our cDNA and genomic human tuftelin structure and sequence submitted to GenBank™ (Accession numbers: AF254260, AF254850, AF254851, AF254852, AF254854, AF254855, AF254856, AF254857, AF254858, AF254859, AF254860, AH009496) with that of the Human Genomic Project data (Accession number: AL365436, released in March 2001) shows that the sequences of the DNA regions corresponding to each of the 13 exons are identical. There is generally a very good agreement with our original estimated sizes of the introns. Three genomic DNA fragments containing the tuftelin genomic DNA sequence, as our results show, were published by the Human Genome Project. Fragment 00104 contained

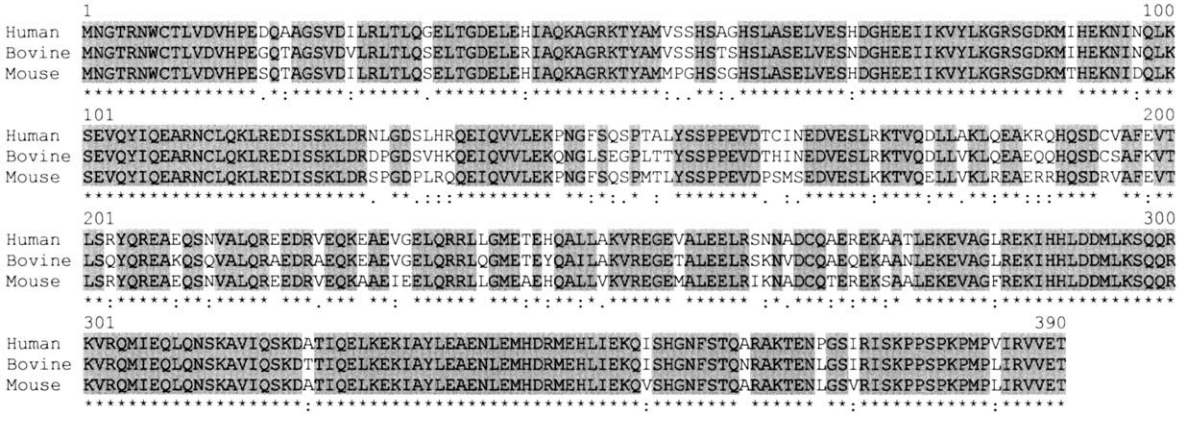

Fig. 8. A comparison between the human, bovine and mouse tuftelin protein sequences. The protein sequences of the human, bovine and mouse tuftelins were derived from the corresponding cDNA sequences. The protein sequences were aligned using the CLUSTAL W Multiple Sequence Alignment Program (Version 1.7). Identical residues are shaded and marked with (*), conserved substitutions are marked with (:) and semi-conserved substitutions are marked with (.) (see Section 3.7).

Table 4
Predicted post-translation modifications of the tuftelin protein

	Human tuftelin	Bovine tuftelin	Mouse tuftelin
N-glycosylation sites	2–5 NGTR	2–5 NGTR	2–5 NGTR
	356–359 NFST	356–359 NFST	356–359 NFST
O-glycosylation sites	152 T	153 T 154 T	153 T
	156 S	156 S	156 S
	157 S	157 S	157 S
	374 S	374 S	374 S
	378 S	378 S	378 S
Phosphorylation sites			
PKC sites	121–123 SSK	121–123 SSK	121–123 SSK
	171–173 SLR	171–173 SLR	171–173 SLK
			192–194 SDR
			269–271 TER
	370–372 SIR	370–372 SIR	370–372 SVR
	378–380 SPK	378–380 SPK	378–380 SPK
CKII sites	9–12 TLVD	9–12 TLVD	9–12 TLVD
	35–38 TGDE	35–38 TGDE	35–38 TGDE
	122–125 SKLD	122–125 SKLD	122–125 SKLD
			127–130 SPGD
	157–160 SPPE	157–160 SPPE	157–160 SPPE
			164–167 SMSE
	175–178 TVQD	175–178 TVQD	175–178 TVQE
		255–258 TALE	269–272 TERE
	322–325 TIQE	322–325 TIQE	322–325 TIQE
EF-hand Ca^{2+} binding site	125–137	125–137	125–137
N-myristoylation sites		17–22 GQTAGS	
	355–360 GNFSTQ	355–360 GNFSTQ	355–360 GNFSTQ
Amidation site	46–49 AGRK	46–49 AGRK	46–49 AGRK

the promoter, exon 1 and part of intron 1, fragment 00723 contained part of intron 1, exons 2–11 and part of intron 11, and fragment 01857 contained part of intron 11, exon 12, intron 12, exon 13 and the 3′ flanking genomic DNA sequence. The alignment of these three fragments (00104, 00723, and 01857) in the correct direction and correct order, fragment 00104 followed by fragment 00723 followed by fragment 01857, will resume the correct human tuftelin gene structure.

RT-PCR of mouse kidney, liver and lung RNA followed by cDNA sequencing revealed that in addition to the tuftelin mRNA transcripts, which contain all 13 exons (liver and lung) (assuming that the mouse tuftelin gene exon–intron organization is the same as in human and bovine), an alternatively spliced mouse tuftelin mRNA transcript missing exon 2 was identified in kidney. Finally, an alternatively spliced mouse tuftelin mRNA lacking exon 2, part of exon 4, exons 5–11 and part of exon 12 was identified in liver. The latter transcript codes for a different protein, which might have another function. This protein consists of 104 amino acids with a MW of 11,468 Da and a basic pI of 9.2. Perhaps the change in protein structure is associated with the change in protein function to suit the specific role of tuftelin in each specific tissue.

The previous data on the bovine tuftelin gene (Bashir et al., 1997) included neither the transcription initiation site nor the structure of the promoter region. In the present study, primer extension analyses using human lung total RNA, corroborated by RT-PCR and DNA sequencing, revealed multiple transcription initiation sites. These transcription initiation sites lacked upstream related TATA boxes. As described above, the most proximal transcription initiation site 93 bp upstream to the ATG start codon was assigned as +1. 5′ RACE studies using human tooth bud total RNA indicated a possible transcription initiation site localized 62 bp upstream to ATG. This site is preceded by an upstream related TATA box and a reverse CAAT box, and promoter prediction software predicted this transcription initiation site with a perfect score. However, no apparent transcription initiation site was identified by the primer extension studies at 62 bp upstream to ATG. Therefore, at least from the primer extension results, there is still no definite evidence to support the 5′ RACE predicted site. The apparent difference between the primer extension results obtained using human lung total RNA and those obtained by 5′ RACE from tooth bud total RNA might be explained by suggesting that the human lung tuftelin mRNA might have different transcription initiation sites to those identified in human tooth bud. The situation is, however, not clear, and future studies on the characterization of the human tuftelin untranslated cDNA region in different tissues might clarify this point. Our very recent RT-PCR followed by DNA sequencing using human tooth bud total RNA revealed the existence of some tuftelin mRNA transcripts with a 5′ untranslated region at least as long as 167 bp upstream to the start codon ATG (i.e. longer than that indicated by the

earlier 5′ RACE results). One problem is obtaining enough human tooth bud total RNA to carry out primer extension studies. The primer extension, RT-PCR, and DNA sequencing results using the human lung total RNA indicate the possibility of some tuftelin mRNA transcripts possessing untranslated regions as long as 796 bp. This is not uncommon in vertebrates (Fan et al., 1995) and might be necessary for efficient initiation of transcription (Kozak, 1991).

The cDNA-derived human protein contains 390 aa, is acidic, hydrophilic, and rich in glutamic acid and aspartic acid. It has a calculated peptide molecular mass of 44,264 Da. Three of the five cysteines found in the human tuftelin are also conserved in the bovine and mouse tuftelin. The presence of cysteines (five in the human, three in the bovine and three in the mouse tuftelins) might suggest the existence of at least one disulfide bond. These are generally important in stabilizing the structure of extracellular proteins. Computer analysis of the post-translation modification of the predicted human tuftelin protein reveals two potential N-glycosylation sites and five potential O-linked glycosylation sites (Table 4). Modification of these residues could greatly affect the protein MW, physical structure and function. In addition, it revealed serine and threonine phosphorylation consensus sites recognized by CKII and PKC, ten in the human, 11 in bovine and 15 in mouse tuftelin (see Table 4).

Ten phosphorylation sites are conserved in all three species. These provide potential sites for the specific chelation of calcium ions and thus could be important in understanding the possible role of tuftelin in enamel biomineralization. These phosphorylation sites are located mainly towards the N-terminal region. A consensus sequence for an EF-hand calcium binding domain was also identified in the human, bovine and mouse tuftelins (with 91, 94 and 93% similarity, respectively). It is located towards the N-terminal region of the tuftelin protein (residues 125–137). One of the extracellular proteins, osteonectin, which is prominent in mineralized tissues, also contains this sequence.

Experiments with the yeast two hybrid system (Paine et al., 1996, 1998, 2000) have demonstrated that tuftelin can self-assemble and that the amino acid residues responsible for self-assembly are present towards the C-terminal region of tuftelin. Perhaps this two-domain structure possessing an anionic domain towards the N-terminal and a self-assembly region towards the C-terminal is important in the function of tuftelin during amelogenesis (Paine et al., 1996, 2000; Deutsch et al., 1997).

Comparison of the predicted amino acid sequences revealed that the human tuftelin shares a very high identity with its bovine and mouse homologs: 89 and 88%, respectively. In addition, immunohistochemistry studies have shown that antibodies against tuftelin reacted with human, bovine, mouse and porcine developing enamel, with the outer layer of pharyngeal teeth of the Pacific hagfish (Vahadi et al., 1996), with the teeth of the lungfish (Satchell et al., 2000) and with shark teeth (Deutsch et al., 1991b). The latter three comprise ancient vertebrate species. High conservation of tuftelin throughout 550 million years of vertebrate evolution suggests that the tuftelin protein has an important biological role. Employing fluorescent *in situ* hybridization, the human tuftelin gene was mapped to chromosome 1q 21-31 (Deutsch et al., 1994). The mapping of the human tuftelin gene to a well-defined cytogenetic region could be important in understanding the etiology of autosomally inherited tooth diseases.

Recently, using tuftelin as a bait protein, Paine et al. (2000) identified a novel tuftelin interacting protein TIP39. They showed that the critical region in the mouse tuftelin protein involved in the interaction with TIP39 was located towards the C-terminal of the tuftelin protein between amino acids 294 and 348 in the mouse tuftelin (MacDougall et al., 1998), and by extrapolation to amino acids 294–348 in the human tuftelin sequence. Paine et al. (1996, 1998, 2000) suggested that this critical region in the tuftelin protein is the same region responsible for tuftelin self-assembly. An amino acid sequence comparison between human, bovine and mouse tuftelins revealed this region to be highly conserved. In addition, Paine et al. (2000) suggested that the TIP39 interaction with tuftelin is related to its tertiary or quaternary structure rather than its secondary structure. Both tuftelin and TIP39 have predicted coiled-coil regions which tend to associate with like domains. They suggested that this may indicate why tuftelin and TIP39 have assembly and self-assembly properties (Paine et al., 1996, 1998) and could also explain the interaction between the two proteins.

Immunocytochemical localization of tuftelin and TIP39 in cultured ameloblast-like cells showed that these two proteins co-localized. Within the developing tooth, the two proteins co-localized in the ameloblast cells, mainly in the Tome's processes of the secretory ameloblast and in the newly secreted extracellular enamel matrix (Paine et al., 2000).

These results regarding the distribution and localization of tuftelin confirm the results of earlier high resolution protein-A gold immunocytochemistry studies (Deutsch et al., 1995) and immunohistochemical studies (Deutsch et al., 1991a; Zeichner-David et al., 1995, 1997; Diekwisch et al., 1997). All these studies showed that tuftelin is expressed in the ameloblast, localized mainly in the Tome's processes (from which enamel extracellular matrix proteins are secreted), is present in extracellular developing enamel and is concentrated at the DEJ. Future isolation and sequencing of tuftelin (protein fragments and peptides) extracted from extracellular enamel will further support the above findings.

Expression of tuftelin and TIP39 starts at a very early stage in odontogenesis (E13 and E14, bud stage in the mouse) (Zeichner-David et al., 1997; Paine et al., 2000). Although the tuftelin protein sequence expressed by the ameloblast does not contain a classical signal peptide sequence (and for that matter neither does the TIP39), it has been found by a number of laboratories to be present in the developing extracellular enamel matrix (Deutsch et al., 1991a, 1995; Zeichner-David et al., 1995; Diekwisch et

al., 1997; Paine et al., 2000). The presence of tuftelin protein in developing extracellular enamel matrix, although it lacks the signal peptide, has been explained by Paine et al. (2000) in two ways: during secretion the ameloblasts seem to shed part of the Tome's processes leaving behind some of its contents (Goldberg et al., 1996, 1999), and secondly the tuftelin-TIP39 complex may play a role in attaching ameloblasts to DEJ or enamel matrix itself; such a protein complex may be shed during secretory activity, thus being incorporated into enamel matrix. Both these explanations are based on the reasoning that tuftelin protein, because it lacks a signal peptide, is probably an intracellular protein. This might be so. It can, however, be argued that perhaps a signal peptide-like motif could be present not at the N-terminal but in the protein interior or C-terminal. Alternatively, it could belong to a small class of proteins that are secreted extracellularly but do not possess a classical signal peptide (for example, see fibroblast growth factors (FGF-9 and FGF-16); Miyake et al., 1998; Revest et al., 2000). It could possibly bind to another protein or complex of proteins which have all the machinery to be secreted.

Whether the tuftelin protein is an intracellular protein or a true extracellular protein because it is (a) expressed at a very early stage of odontogenesis (bud stage) (where ectodermal/mesenchymal interactions are taking place to eventually produce the ameloblasts and their secretory extracellular enamel matrix, and the odontoblasts and their secretory extracellular dentin matrix), (b) expressed in the secretory ameloblast cells and localized in the Tome's process (through which extracellular enamel matrix proteins are secreted) and (c) concentrates at the DEJ region (where enamel formation and initial mineralization commence), it could exert considerable influence on enamel protein secretion, assembly and on the initial enamel mineralization at the DEJ, thus playing an important role in odontogenesis. It might have a role in cell transportation and/or secretion, as suggested by Paine et al. (2000).

Searching the human EST database (Table 5) revealed the expression of partial tuftelin protein coding cDNA sequences in human adult testis, adult female breast, kidney renal cell adenocarcinoma, lung small cell carcinoma and placenta choriocarcinoma. In addition, cDNA sequence identity to the corresponding 3′ untranslated region of the human tuftelin cDNA (exon 13) (it is assumed that sequencing larger cDNA fragments 5′ to the present available 3′ untranslated sequences would have revealed the full translatable human tuftelin cDNA) was found in germ-cell tumor, ovary tumor, colon, stomach poorly differentiated adenocarcinoma, lung squamous cell carcinoma, lung carcinoid, uterus well-differentiated endometrial adenocarcinoma, soares senescent fibroblasts, and promyelocyte. A search in the mouse EST database (Table 6) revealed, in addition to the above observations, that tuftelin was also expressed in mouse embryonic stem (ES) cells, skin, thymus, vagina, pituitary gland, hypothalamus, tongue, colon, pancreas islet, eyeball, brain, medulla oblongata

Table 5
ESTs matching the human tuftelin cDNA sequence

Tissues	GenBank Accession numbers (corresponding human tuftelin cDNA region)
ESTs containing partial protein coding sequences	
Adult testis	AL043356 (846–1578)*[a]
Adult female breast	BE074742 (468–717)
Kidney renal cell adenocarcinoma	BE273276 (70–701)*
Lung small cell carcinoma	BE263406 (69–154 and 228–693)*
Placenta choriocarcinoma	BE903227 (72–154, 228–853)*
	BE733924 (63–154, 228–882)*
ESTs containing partial 3′ untranslated region	
Adult testis	AI187417 (2723–3139)
Adult female breast	AA527170 (2614–3139)
Germ cell tumor	AI350867 (2759–3133)
Ovary tumor	AA400960 (2593–2997)
	AA430389 (2706–3131)
	AA400891 (2811–3131)
Colon	AA534506 (2723–3132)
Stomach poorly differentiated adenocarcinoma with signet ring cell features	AW131222 (2783–3138)
	AI925461 (2678–3133)
Lung two pooled squamous cell carcinomas	AW264729 (2722–3133)
Lung carcinoid	AI458334 (2816–3126)
Uterus, well-differentiated endometrial adenocarcinoma, pooled tumors	AI699870 (2754–3133)
Soares senescent fibroblasts	W51933 (2662–3127)
	AA284284 (2662–3108)
	W48637 (2751–3132)
Promyelocyte	D20078 (2729–3131)

[a] ESTs contain partial protein coding sequence and partial untranslated region.

Table 6
ESTs matching the mouse tuftelin cDNA sequence

Tissues	GenBank Accession numbers (corresponding mouse tuftelin cDNA region)	
ESTs containing partial protein coding sequence and partial 3′ untranslated region		
Skin	AA611681 (1185–1641)	AA612327 (1185–1553)
Mammary tumor	BE911159 (1051–1755)	BE913991 (1129–1491)
ESTs containing 3′ untranslated region		
ES cells	AV215660 (2275–2569)	
Embryo	AU044078 (2148–2569)	AA612052 (2027–2448)
	BB377326 (2333–2572)	AU045470 (2051–2569)
	BB416752 (2304–2570)	BB045069 (2314–2572)
	BB527032 (2394–2572)	AV296094 (2301–2569)
	BB438958 (2291–2570)	BB408204 (2278–2570)
	BB527066 (2342–2572)	
Embryo stomach	BB492671 (2273–2570)	BB493995 (2309–2570)
Testis	AV276966 (2316–2569)	
Mammary gland	AA616113 (1982–2564)	BB512883 (2379–2572)
Mammary tumor	BE914129 (1900–2566)	BE373023 (1383–2000)
Skin	BB008113 (2309–2569)	AV231101 (2329–2572)
	BB149738 (2345–2569)	AV239787 (2353–2572)
	BB004220 (2406–2572)	AV240399 (2351–2572)
	AV233203 (2336–2569)	AV229142 (2347–2569)
	BB151280 (2353–2572)	AV239069 (2340–2570)
Thymus	AA212596 (1463–1999)	BB030770 (2347–2572)
Vagina	BB146836 (2254–2572)	BB144461 (2318–2570)
	BB148919 (2340–2572)	
Pituitary	BB022603 (2340–2572)	
Hypothalamus	AW121475 (2294–2566)	BE979176 (2404–2567)
Tongue	AV084667 (2377–2571)	AV092689 (2350–2571)
Barstead bowel	AA871221 (2024–2545)	
Colon	AV372719 (2313–2572)	
Cecum	AV373846 (2331–2571)	AV378218 (2359–2571)
Pancreas islet	BB436862 (2280–2572)	
Eyeball	BB545974 (2280–2572)	
Head	AV292078 (2346–2571)	
Brain	BE648900 (1627–2001)	
Medulla oblongata	BB047089 (2341–2572)	
Corpora quadrigemina	BB302612 (2335–2569)	

and corpora quadrigemina. This suggests that tuftelin has a universal and important role in cell function and metabolism. The appearance of tuftelin in ES cells as well as in many different carcinomas is particularly interesting.

Our recent preliminary indirect immunohistochemistry studies (Deutsch et al., unpublished data) have shown that in addition to the expression of tuftelin mRNA the protein itself is also expressed in different non-mineralizing tissues.

The present results, and our ongoing studies on the production of a tuftelin knock-out mouse, will provide an important base-line for future understanding of the biological role of tuftelin.

Acknowledgements

This work was supported by the US–Israel Binational Science Foundation Grant Number 95-00347. D. Deutsch is a member of the Minerva Center for Calcium and Bone Metabolism.

References

Bashir, M.M., Abrams, W.R., Rosenbloom, J., 1997. Molecular cloning and characterization of the bovine tuftelin gene. Arch. Oral Biol. 42, 489–496.

Bashir, M.M., Abrams, W.R., Tuker, T., Sellinger, B., Budarf, M., Emanuel, B., Rosenbloom, J., 1998. Molecular cloning and characterization of the bovine and human tuftelin genes, Connect. Tissue Res., 39, 13–24, discussion 63–67.

Catalano-Sherman, J., Palmon, A., Burstein, Y., Deutsch, D., 1993. Amino acid sequence of a major human amelogenin protein employing Edman degradation and cDNA sequencing. J. Dent. Res. 72, 1566–1572.

Deutsch, D., Palmon, A., Fisher, L.W., Termine, J., Young, M.F., 1989. Cloning of bovine enamelin. In: Fearnhead, R.W. (Ed.). Tooth Enamel, V. Florence, Yokohama, pp. 308–313.

Deutsch, D., Palmon, A., Fisher, L.W., Kolodny, N., Termine, J.D., Young, M.F., 1991a. Sequencing of bovine enamelin (“tuftelin”) – a novel acidic enamel protein. J. Biol. Chem. 266, 16021–16028.

Deutsch, D., Palmon, A., Dafni, L., Shenkman, A., Sherman, J., Fisher, L.W., Termine, J.D., Young, M.F., 1991b. Enamelin and enameloid. In: Suga, S., Nakahara, H. (Eds.). Mechanisms and Phylogeny of Mineralization in Biological System, Springer-Verlag, Tokyo, pp. 73–77.

Deutsch, D., Palmon, A., Young, M.F., Selig, S., Kearns, W.G., Fisher,

L.W., 1994. Mapping of the human tuftelin (TUFT1) gene to chromosome 1 by fluorescence in situ hybridization. Mamm. Genome 5, 461–462.

Deutsch, D., Palmon, A., Dafni, L., Catalano-Sherman, J., Young, M.F., Fisher, L.W., 1995. The enamelin (tuftelin) gene. Int. Dev. Biol. 39, 135–143.

Deutsch, D., Dafni, L., Palmon, A., Hekmati, M., Young, M.F., Fisher, L.W., 1997. Tuftelin: enamel mineralization and amelogenesis imperfecta (review). In: Chadwick, D.J., Cardew, G. (Eds.). Dental Enamel, Ciba Foundation Symposium, 205. Wiley, Chichester, pp. 135–155.

Deutsch, D., Palmon, A., Dafni, L., Mao, Z., Leytin, V., Young, M., Fisher, L.W., 1998. Tuftelin – aspects of protein and gene structure. Eur. J. Oral Sci. 106 (Suppl. 1), 315–323.

Diekwisch, T.G., Ware, J., Fincham, A.G., Zeichner-David, M., 1997. Immunohistochemical similarities and differences between amelogenin and tuftelin gene products during tooth development. J. Histochem. Cytochem. 45, 859–866.

Fan, N.C., Peng, C., Krisinger, J., Leung, P.C., 1995. The human gonadotropin-releasing hormone receptor gene: complete structure including multiple promoters, transcription initiation sites, and polyadenylation signals. Mol. Cell. Endocrinol. 107, R1–R8.

Frohman, M.A., Dush, M.K., Martin, G.R., 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85, 8998–9002.

Goldberg, M., LeColle, S., Bissila-Mapahou, P., Septier, D., Carreau, J.P., 1996. Radioautographic study of the incorporation of (3H)-choline into the phospholipids of secretory ameloblasts and enamel of normal and essential-fatty-acid-deficient rats. Adv. Dent. Res. 10, 126–134.

Goldberg, M., Lecolle, S., Vermelin, L., Benghezal, A., Godeau, G., 1999. [^{3}H]choline uptake and turnover into membrane and extracellular matrix phospholipids, visualized by radioautography in rat incisor dentin and enamel. Calcif. Tissue Int. 65, 66–72.

Kozak, M., 1991. Effects of long 5′ leader sequences on initiation by eukaryotic ribosomes in vitro. Gene Exp. 1, 117–125.

MacDougall, M., Simmons, D., Dodds, A., Knight, C., Luan, X., Zeichner-David, M., Zhang, C., Ryu, O.H., Qian, Q., Simmer, J.P., Hu, C.C., 1998. Cloning, characterization, and tissue expression pattern of mouse tuftelin cDNA. J. Dent. Res. 77, 1970–1978.

Miyake, A., Konishi, M., Martin, F.H., Hernday, N.A., Ozaki, K., Yamamoto, S., Mikami, T., Arakawa, T., Itoh, N., 1998. Structure and expression of a novel member, FGF-16, on the fibroblast growth factor family. Biochem. Biophys. Res. Commun. 243, 148–152.

Paine, M.L., Deutsch, D., Snead, M.L., 1996. Carboxyl-region of tuftelin mediates self-assembly. Connect. Tissue Res. 35, 157–161.

Paine, M.L., Krebsbach, P.H., Chen, L.S., Paine, C.T., Yamada, Y., Deutsch, D., Snead, M.L., 1998. Protein-to-protein interactions: criteria defining the assembly of the enamel organic matrix. J. Dent. Res. 77, 496–502.

Paine, C.T., Paine, M.L., Luo, W., Okamoto, C.T., Lyngstadaas, S.P., Snead, M.L., 2000. A tuftelin-interacting protein (TIP39) localizes to the apical secretory pole of mouse ameloblasts. J. Biol. Chem. 275, 22284–22292.

Revest, J.M., DeMoerlooze, L., Dickson, C., 2000. Fibroblast growth factor 9 secretion is mediated by a non-cleaved amino-terminal signal sequence. J. Biol. Chem. 275, 8083–8090.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, . 2nd EditionCold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Satchell, P.G., Shuler, C.F., Diekwisch, T.G., 2000. True enamel covering in teeth of the Australian lungfish Neoceratodus forsteri. Cell Tissue Res. 299, 27–37.

Vahadi, R., Marches, F., Diekwisch, T.G.H., 1996. Evidence for enamel protein gene expression in cyclostome pharyngeal teeth. J. Dent. Res. 75, 76.

Zeichner-David, M., Diekwisch, T., Fincham, A., Lau, E., MacDougall, M., Moradian-Oldak, J., Simmer, J., Snead, M., Slavkin, H.C., 1995. Control of ameloblast differentiation. Int. J. Dev. Biol. 39, 69–92.

Zeichner-David, M., Vo, H., Tan, H., Diekwisch, T., Berman, B., Thiemann, F., Alcocer, M.D., Hsu, P., Wang, T., Eyna, J., Caton, J., Slavkin, H.C., MacDougall, M., 1997. Timing of the expression of enamel gene products during mouse tooth development. Int. J. Dev. Biol. 41, 27–38.